Foata's Bijection for Tree-Like Structures

Oleg Pikhurko*
DPMMS, Centre for Mathematical Sciences
Cambridge University, Cambridge CB3 0WB, England
E-mail: O.Pikhurko@dpmms.cam.ac.uk

February 1, 2008

Abstract

We present bijections enumerating (k, m)-trees, k-gon trees, edge labelled $(2, 1)$-trees, and other tree-like structures. Our constructions are based on Foata's [8] bijection for cycle-free functions, which is simplified here.

1 Introduction

Here we enumerate certain tree-like structures by finding an explicit bijection between the set in question and some trivially simple set. Although some of the obtained formulas are known, the constructed bijections are new results. Bijective proofs are of interest as they carry more information than formulas alone: bijections allow us to generate all objects one by one and it is often possible to enumerate some subclasses by analysing the corresponding codes.

For example, the analytic formula for the number of vertex-labelled $(m + 1, m)$-trees was independently found by Beineke and Pippert [1] and Moon [11]. (We will provide all definitions in due course.) Nevertheless, a number of different bijective proofs of the result appeared as well, see [14, 8, 9, 7, 5].

One of these papers, by Foata [8], contains an interesting bijection for cycle-free functions. In Section 2 we describe Foata's construction with a simpler argument than the original one.

Cycle-free functions provide a natural framework for encoding different tree-like structures, in particular $(m + 1, m)$-trees, as demonstrated by

*Supported by a Senior Rouse Ball Studentship, Trinity College, Cambridge, UK.

Foata [8]. Extending Foata's argument, we present a bijection for general (k, m)-trees. (The analytic formula was found by the author [13].)

This method can enumerate some other objects. For example, we can find a bijection for vertex labelled *k-gon trees* (also known as *cacti* or *trees of polygons*), a structure that appears in [4, 15, 12, 10]. The derived formula for the number of vertex-labelled k-gon trees seems to be a new result.

In order not to repeat the same portions of proof twice, we enumerate a more general structure, *H-built-trees*, which includes (k, m)-trees and k-gon trees as partial cases. Please refer to Section 3 for details.

In Section 4 we present a bijection for edge labelled $(2, 1)$-trees. This answers a question posed by Cameron [3] which was motivated by the possibility that such a bijection can simplify some of his enumeration results (or proofs). However, although we answered Cameron's question, we were not able to improve on [3].

Unfortunately, we do not know any bijection (nor even a compact analytic formula) giving the number of edge labelled (k, m)-trees for $k \geq 3$.

2 Foata's Bijection

Given disjoint finite sets A, B, C and a surjection $\gamma : B \to A$, a function $f : A \to B \cup C$ is called *cycle-free* if for every $b \in B$ the sequence $(f \circ \gamma)^i(b)$ eventually terminates at some $c \in C$.

Foata [8, Theorem 1] exhibited a bijection between $F(A, B, C, \gamma)$, the set of cycle-free functions, and the set of functions $g : A \to B \cup C$ such that $g(a_1) \in C$ for some beforehand fixed $a_1 \in A$; this implies

$$|F(A, B, C, \gamma)| = |C|(|B| + |C|)^{|A|-1}. \tag{1}$$

We briefly describe a simpler construction than that in [8]. Fix some ordering of A. Let $f \in F(A, B, C, \gamma)$. Let $Z = (z_1, \ldots, z_s)$ denote the increasing sequence of the elements in $A \setminus \gamma(f(A))$. (For convenience we assume that $\gamma(c) = c$ for $c \in C$.) We build, one by one, s sequences $\delta_1, \ldots, \delta_s$ composed of elements in $B \cup C$. Having constructed sequences $\delta_1, \ldots, \delta_{i-1}$, let $m_i \geq 0$ be the smallest integer such that $(\gamma \circ f)^{m_i+1}(z_i)$ either belongs to C or occurs in at least one of $\gamma \otimes \delta_1, \ldots, \gamma \otimes \delta_{i-1}$, where $\gamma \otimes (x_1, \ldots, x_i) \equiv (\gamma(x_1), \ldots, \gamma(x_i))$. We define (mind the order)

$$\delta_i = \Big((f \circ \gamma)^{m_i}(f(z_i)),\ (f \circ \gamma)^{m_i-1}(f(z_i)),\ \ldots,\ f(z_i)\Big). \tag{2}$$

One can easily check that Z is non-empty if A is, every m_i exists, and δ, the juxtaposition product of the s sequences $\delta_1, \ldots, \delta_s$, contains $|A|$ elements.

(In fact, δ is but a permutation of $(f(a))_{a\in A}$.) The obtained sequence δ of $|A|$ elements of $B\cup C$, which starts with an element of C, corresponds naturally to the required function $g: A \to B\cup C$.

Conversely, given g (or δ) we can reconstruct $Z = A \setminus \gamma(g(A))$. Then, exactly $s = |Z|$ times, an element of $\gamma\otimes\delta$ either belongs to C or equals some preceding element. These s positions mark the beginnings of $\delta_1, \dots, \delta_s$. Now we can restore the required f by (2). To establish (1) completely, one has to check easy details.

3 H-Built-Trees

The following notion of *(k,m)-tree* was suggested independently by Dewdney [6] and Beineke and Pippert [2].

Let us agree that the vertex set is $[n] \equiv \{1, \dots, n\}$. Fix the *edge size* k and the *overlap size* $m \in [0, k-1] \equiv \{0, \dots, k-1\}$. We refer to k-subsets and m-subsets of $[n]$ as *edges* and *laps* respectively. A non-empty k-graph (i.e. a k-uniform set system) without isolated vertices is called a *(k,m)-tree* if we can order its edges, say $E_1, \dots, E_e$, so that, for every $i \in [2, e]$, there is $j \in [i-1]$ such that $|E_i \cap E_j| = m$ and $(E_i \setminus E_j) \cap \left(\cup_{h=1}^{i-1} E_h\right) = \emptyset$. In other words, we start with a single edge and can consecutively affix a new edge along an m-subset of an existing edge. For example, a $(k,0)$-tree consists of disjoint edges; $(2,1)$-trees are usual (Cayley) trees.

Adopting the ideas of Foata [8], we present a bijection for (k,m)-trees. In fact, we enumerate a more general structure defined as follows.

Let H be any m-graph on $[k]$. An *H-built-tree* $(T, \{H_1, \dots, H_e\})$ consists of a usual (k,m)-tree T with edges $E_1, \dots, E_e$ plus H-graphs H_i on E_i, $i \in [e]$, such that if $E_i \cap E_j$ is a lap (that is, has size m), then it is an edge of both H_i and H_j for any distinct $i, j \in [e]$. Let $n = e(k-m)+m$ denote the total number of vertices and let

$$f = \left|\cup_{i\in[e]} E(H_i)\right| = e(l-1)+1,$$

where $l = e(H) = |E(H)|$ denotes the number of edges in H. Also, let $\mathcal{R}_H$ be the set of distinct H-graphs on $[k]$ rooted at $[m]$, that is, containing $[m]$ as an edge. Clearly,

$$|\mathcal{R}_H| = \frac{k!\,l}{|\mathrm{Aut}(H)|\binom{k}{m}}.$$

An H-built-tree is *rooted* on an m-set L if $L \in \cup_{i\in[e]} E(H_i)$.

Theorem 1 *There is a bijection between the set Y of H-built-trees on $[n]$ rooted at $[m]$ and the set*

$$Z = F(A,B,C,\gamma) \times \prod_{i=1}^{e} (X_i \times \mathcal{R}_H)\,,$$

where $A = [e]$, $B = [e] \times [l-1]$, $C = \{[m]\}$, γ is the coordinate projection $B \to A$, and $X_i = [\binom{(k-m)(e-i+1)-1}{k-m-1}]$. In particular,

$$\begin{aligned} |Y| &= f^{e-1}\left(\frac{k!\,l}{\binom{k}{m}|\mathrm{Aut}(H)|}\right)^{e} \prod_{i=1}^{e} \binom{(k-m)(e-i+1)-1}{k-m-1} \\ &= \frac{(e(k-m))!\,f^{e-1}}{e!}\left(\frac{m!\,l}{|\mathrm{Aut}(H)|}\right)^{e}. \end{aligned}$$

Proof. Given an H-built-tree T rooted at $[m]$, order its edges $E_1, \dots, E_e$ so that $[m] \in E(H_1)$ and each E_i, $i \in [2,e]$, shares a lap with some E_j, $j < i$.

Correspond an edge E_i to the lap $g'(E_i) = E_i \cap \cup_{j=1}^{i-1} E_j$, $i \in [2,e]$. (We agree that $g'(E_1) = [m]$.) Call the set $f(E_i) = E_i \setminus g'(E_i)$ the *free part* of E_i; the free parts partition $[n] \setminus [m]$. Clearly, these definitions of g' and f do not depend on the particular ordering.

Relabel the edges by $D_1, \dots, D_e$ so that $d_i = \min f(D_i)$ increases; let H'_i denote the corresponding H-graph on D_i. Label, in the colex order, all edges (laps) of H'_i but $g'(D_i) \in E(H'_i)$ by (i,j), $j = 1, \dots, l-1$. Note that now we have indexing of the edges of T by A, namely $(D_i)_{i \in A}$, and of the laps of T by $B \cup C$. Let $g : A \to B \cup C$ be the map corresponding to g'. A moment's thought reveals that g is cycle-free.

Repeat the following for $i = 1, \dots, e$. Index, in the colex order, the $(k-m-1)$-subsets of $(\cup_{j=i}^{e} f(D_j)) \setminus \{d_i\}$ by the elements of X_i and let $x_i \in X_i$ be the index corresponding to $f(D_i) \setminus \{d_i\}$. Consider the bijection $h : D_i \to [k]$ such that h is monotone on $g'(D_i)$ and $f(D_i)$ which are respectively mapped onto $[m]$ and $[m+1,k]$. Let $R_i \in \mathcal{R}_H$ be the image of H'_i under h.

Now, $(g, x_1, R_1, \dots, x_e, R_e) \in Z$ is the 'code' of $T \in Y$.

Conversely, given an element $(g, x_1, R_1, \dots, x_e, R_e) \in Z$ we can consecutively reconstruct the sequence $(d_i, f(D_i))$, $i = 1, \dots, e$. Indeed, d_i is the smallest element of $V = [n] \setminus ((\cup_{j=1}^{i-1} f(D_j)) \cup [m])$ while $f(D_i) \setminus \{d_i\}$ is the x_ith $(k-m-1)$-subset of $V \setminus \{d_i\}$. For $i \in A$ with $g(i) \in C$, we have $D_i = [m] \cup f(D_i)$ and (knowing $g'(D_i) = [m]$ and $f(D_i)$), we can determine H'_i from R_i; then we can recover the lap corresponding to $(i,j) \in B$ as the jth element of $E(H'_i) \setminus \{[m]\}$, $j \in [l-1]$.

Likewise, we can reconstruct all information about D_i for any $i \in A$ with $g(i)$ being already associated with a lap. As f is cycle-free, all edges are eventually identified, producing $T \in Y$.

The plain verification shows that we have indeed a bijective correspondence between Y and Z. ∎

It is easy to see that (k,m)-trees bijectively correspond to K_k^m-built-trees on the same vertex set, where K_k^m denotes the complete m-graph of order k. Now, $n = e(k-m)+m$, $l = \binom{k}{m}$, $f = e(l-1)+1$, $|\mathcal{R}_{K_K^m}| = 1$, and we deduce the following.

Corollary 2 *The number of (k,m)-trees on $[n]$ rooted at $[m]$ equals*

$$R_{km}(e) = \frac{(n-m)!\,f^{e-1}}{e!\,((k-m)!)^e}. \ ∎ \tag{3}$$

Remark. Clearly, the number of vertex labelled (k,m)-trees with n vertices is equal to $\binom{n}{m} R_{km}(e)/f$, which gives precisely [13, formula (1)].

As another consequence, we can enumerate *k-gon trees*, which are inductively defined as follows. A k-gon (that is, C_k, a cycle of length k) is a k-gon tree. A k-gon tree with $e+1$ k-gons is obtained from a k-gon tree with e k-gons by adding $k-2$ new vertices and a new k-gon through these vertices and an already existing edge.

It is trivial to check that if a union of C_k-graphs can be formed into a C_k-built-tree, then the latter is uniquely defined. Thus, k-gon trees are in bijective correspondence with C_k-built-trees. We have $m = 2$, $n = e(k-2)+2$, $f = e(k-1)+1$, and $|\mathcal{R}_{C_k}| = (k-2)!$, so we obtain that there are $(e(k-2))!\,f^{e-1}/e!$ rooted k-gon trees with e k-gons, which implies the following result.

Corollary 3 *The number of vertex labelled k-gon trees with e k-gons is*

$$\frac{(e(k-2)+2)!\,(e(k-1)+1)^{e-2}}{2(e!)}, \quad k \geq 3. \ ∎$$

4 Edge Labelled Trees

Cameron [3] enumerates certain classes of what is called there *two-graphs*: reduced, 5-free, and $(5,6)$-free. Please refer to his work for all definitions and details. Also, he defines, for a given (Cayley) tree T, the equivalence relation $\cong$ on its edges which is the smallest one such that two edges are

related if they intersect at a vertex of degree 2 in T. For example, T is series-reduced if and only if $\cong$ is the identity relation.

Cameron had to count the number S_n of trees with n labelled edges when we do not distinguish trees obtained by permuting labels within the $\cong$-classes. He found the following formula [3, Proposition 3.5(a)]:

$$S_n = \sum_{k=1}^{n} S(n,k) \frac{1}{k+1} \sum_{j=0}^{k-1} (-1)^j \binom{k+1}{j} \binom{k-1}{j} j!(k-j+1)^{k-j-1}, \quad (4)$$

where $S(n,k)$ is the Stirling number of the second kind. The sequence (S_n) starts as $1, 1, 2, 8, 52, \dots$ and probably cannot be represented in a closed expression but, of course, one can try to simplify (4). Cameron [3] asks the following question.

Problem 4 (Cameron) *Describe a constructive bijection for edge labelled trees, not going via vertex labelling. Describe the equivalence relation $\cong$ in terms of this code.*

The motivation for the problem was that such a code might simplify (4). Although we answer here this question, we do not see how our bijection can simplify (4) or its proof from [3].

Let us describe our construction. We use Foata's [8] bijection for cycle-free functions.

Let $e_1, \dots, e_n$ be the edges. Suppose $e_1 = \{a, b\}$; this edge will play a special role. Let $A = B = \{e_2, \dots, e_n\}$, $C = \{a, b\}$ and $\gamma : B \to A$ be the identity function. Let us correspond an $f \in F(A, B, C, \gamma)$ to a given tree T. Each edge e can be connected to e_1 by the unique path in T. If e is incident to e_1, then let $f(e)$ be equal to their common vertex; otherwise, let $f(e)$ be the first edge on the path from e to e_1. This gives a correspondence between twice the number of edge-labelled trees (we can label the two vertices of e_1 by a and b in two different ways) and $F(A, B, C, \gamma)$. Foata's bijection shows that $|F(A, B, C, \gamma)| = |C|(|A| + |C|)^{|A|-1}$, which implies, as desired, that the number of edge-labelled trees with n edges is $(n+1)^{n-2}$.

To make this correspondence one-to-one, we can consider only a half of the codes, e.g. those starting with a. They correspond to trees in which the path from the leaf e_i with the smallest index $i \in [2, n]$ to e_1 hits e_1 at a.

Of course, the code is rather simple; we describe briefly only one direction. A code δ is a sequence of length $n-1$ consisting of elements in $\{a, b, e_2, \dots, e_n\}$ and staring with a. The set $Z \subset \{e_2, \dots, e_n\}$ of edges which do not occur in the sequence consists of leaves. (If b does not occur,

then e_1 is also a leaf.) Clearly, an element of δ equals either a or b or some previously occurring element exactly $s = |Z|$ times. Cut δ before each such element; we have s pieces $\delta_1, \dots, \delta_s$. Append the ith element z_i of Z to the end of δ_i to obtain δ_i', $i \in [s]$.

The reversed sequence δ_i' describes the initial segment P_i' of the path P_i from the element $z_i \in Z$ to e_1 until it hits e_1 or some previous path P_j, $i \in [s]$. Clearly, this determines some tree.

How can we read the $\cong$-relation from δ? First, let $\cong'$ be the minimal equivalence relation on $\{a, b, e_2, \dots, e_n\}$ such that $e_i \cong' e_j$ if e_i and e_j intersect at a vertex of degree 2, $2 \le i < j \le n$, and $x \cong' e_i$ if x is a degree-2 vertex incident to e_i, $x \in \{a, b\}$, $i \in [2, n]$. (Informally, we cut e_1 in the middle and take the usual $\cong$-relation on the both created components separately.) Clearly, $\cong$ is obtained from $\cong'$ by identifying a and b into a single element e_1, so let us indicate how to determine the latter relation.

Take any maximal contiguous subsequence $S \subset \delta$ consisting of elements that occur in δ exactly once. Clearly, S lies entirely within some δ_i and $S \cup \{y\}$ is a $\cong'$-equivalence class, where y is the symbol following S in δ_i'. Conversely, it is easy to check that all non-trivial $\cong'$-classes are obtained this way, as required.

References

[1] L. W. Beineke and R. E. Pippert. The number of labelled k-dimensional trees. *J. Combin. Theory*, 6:200–205, 1969.

[2] L. W. Beineke and R. E. Pippert. On the structure of (m, n)-trees. In *Proc. 8th Southeast Conference on Combinatorics, Graph Th. and Computing (Louisiana State Univ., Baton Rouge)*, volume XIX of *Congressus Numeratium*, pages 75–80. Utilitas Math., Winnipeg, 1977.

[3] P. J. Cameron. Counting two-graphs related to trees. *Electronic J. Combin.*, 2(#R4):8pp, 1995.

[4] C.-Y. Chao and N.-Z. Li. Trees of polygons. *Archiv Math.*, 45:180–185, 1985.

[5] W. Y. C. Chen. A coding algorithm for Rényi trees. *J. Combin. Theory* (A), 63:11–25, 1993.

[6] A. K. Dewdney. Multidimensional tree-like structures. *J. Combin. Theory* (B), 17:160–169, 1974.

[7] O. Egecioglu and L.-P. Shen. A bijective proof for the number of labelled q-trees. *Ars Combinatoria*, 25:3–30, 1988.

[8] D. Foata. Enumerating k-trees. *Discrete Math.*, 1:181–186, 1971.

[9] C. Greene and G. A. Iba. Caley's formula for multidimensional trees. *Discrete Math.*, 14:1–11, 1975.

[10] K. M. Koh and C. P. Teo. Chromaticity of series-parallel graphs. *Discrete Math.*, 154:289–295, 1996.

[11] J. W. Moon. The number of labeled k-trees. *J. Combin. Theory*, 6:196–199, 1969.

[12] Y.-H. Peng. On the chromatic uniqueness of certain trees of polygons. *J. Austral. Math. Soc. Ser. A*, 55:403–410, 1993.

[13] O. Pikhurko. Enumeration of labelled (k, m)-trees. *J. Combin. Theory* (A), 86:197–199, 1999.

[14] C. Rényi and A. Rényi. The Prüfer code for k-trees. In P. Erdős et al., editors, *Combinatorial Theory and Its Applications*, pages 945–971. North-Holland, Amsterdam, 1970.

[15] E. G. Whitehead, Jr. Chromatic polynomials of generalized trees. *Discrete Math.*, 72:391–393, 1988.